DOCUMENT RESUME

ED 242 068 EA 016 590

AUTHOR Patton, Michael Quinn
TITLE Sneetches, Zax and Empty Pants: Alternative Approaches to Evaluation.
PUB DATE Feb 84
NOTE 16p.; Paper presented at the Meeting of the California Postsecondary Conference on Evaluation (February 1984).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Decision Making; Educational Administration; *Educational Assessment; Elementary Secondary Education; *Evaluation; Evaluation Criteria; Evaluation Methods; Evaluators; Information Systems; *Information Utilization; *Program Evaluation; Program Improvement; Remedial Instruction; Summative Evaluation

ABSTRACT

Involving producing, analyzing, and using information, evaluation is the central discipline of the information age. Both traditional criteria for good evaluation and standards formulated in 1981 prescribe evaluations that are useful, practical, ethical, and accurate. The burning question in evaluation is the utilization of information for decision-making. Incidents from stories by Dr. Seuss exemplify three major barriers to the utilization of evaluation information: fear, methods madness, and dogmatism. Educators' fear of judgment can be dealt with by placing evaluation in a cultural context, viewing it as one of the many possible ways of looking at remedial education programs. Methods madness occurs when attention to measurement dominates the evaluation process and methods become an end in themselves. The removal of this barrier lies in the current consensus that whether qualitative or quantitative methods are used, the methods must be assessed in terms of what questions are asked and what kinds of data are needed. Dogmatism in evaluation models and in preconceptions about what evaluation means can be countered by creativity, flexibility, situational responsiveness, and sensitivity to different ways of raising questions. Factors contributing to a high level of utilization of an evaluation exemplar are appropriate and relevant methods, active involvement of people affected, a clear focus to research, and clear action implications specified in advance. (MJL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED242068

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SNEETCHES, ZAX AND EMPTY PANTS: ALTERNATIVE APPROACHES TO EVALUATION

by Michael Quinn Patton
University of Minnesota

February 1984

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Michael Q. Patton
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This chapel strikes me as a particularly appropriate setting for an evaluation conference. A lot of people spend a lot of time praying...

--praying an evaluation won't happen;
--praying that the date for submitting remedial education evaluation plans will be extended indefinitely; and
--praying that evaluation, when it inevitably comes, will somehow be positive.

All and all, evaluation can be a very religious experience. Therefore, I take as my text for tonight the Genesis story of Paradise Lost.

> In the beginning God created the heaven and the earth....
>
> And God saw every thing that He made. "Behold," God said, "it is very good." And the evening and the morning were the sixth day.
>
> And on the seventh day God rested from all His work. His archangel came then unto Him asking: "God, how do you know that what you have created is 'very good'? What are your criteria? On what data do you base your judgment? Aren't you a little close to the situation to make a fair and unbiased evaluation?"
>
> God thought about these questions all day and His rest was greatly disturbed. On the eighth day God said, "Lucifer, go to hell."
>
> Thus was evaluation born in a blaze of glory.
> Ever since the status of the profession has been somewhat in doubt: the road to salvation or a sure ticket to damnation?
> (Patton, 1981:17)

EA 016 590

The Emergence of the Information Age

It seems to me appropiate at a conference like this to begin with the

beginning, and to begin with an overview of how we got to where we are today. Such an overview would include a look back 3 to 5 million years to the beginning of our species on earth. The anthropologists tell us that we have spent most of our existence on earth in small hunting and gathering societies where the primary activity was the struggle for survival. About 10 thousand years ago the age of agriculture began as we learned how to grow our own food. Only 300 years ago the industrial revolution began. Now, in the last 20 to 30 years, we have emerged into "Post-Industrial Society," "The Third Wave," or "The Information Age." The significance of the evolution into post-industrial society is that for the first time in our history a majority of the population is engaged in activities that do not involve physically producing something. In the pre-agricultural and agricultural periods the search for food predominated. During the industrial period the manufacture of goods from raw materials predominated. In post-industrial society the great majority of people are involved in service activities, manipulating information, generating knowledge, and applications of knowledge through management, teaching, and the whole host of activities that characterize the information age.

Evaluation epitomizes the information age. Evaluation involves producing, analyzing, and using information. Remedial education is also a primary symbol of what the information age means. Remedial education symbolizes the importance we attach to the abilities necessary for acquiring and using information through reading, manipulating numbers, and functional literacy. People who lack such skills are highly disadvantaged in the information age. Professionals who lack evaluation skills are also highly disadvantaged in the information age.

Central Problems and Burning Questions

The central problem of the agricultural period was the production of food

and the key resource of that period was land. The central problem of the industrial age was converting raw resources into manufactured goods, and the most important resource was capital. The central problem of the information age is obtaining and applying useful knowledge, and the coin of the realm is information. Never has it been more true that knowledge is power.

The utilization of information is the key problem of the information age. The challenge is not really producing information or storing information but rather getting people to use information. In such fields as medicine, nutrition, education, agriculture, and other fields of knowledge, technological advances have moved far ahead of the willingness of human beings to apply new knowledge. Compliance studies in the field of medicine indicate widespread disregard of doctors' advice. Agricultural extension agents and health educators face substantial obstacles in attempting to get people to apply knowledge we already have. The problem of appropriate information use is generic to the information age.

Each discipline is characterized by a burning question. In sociology the burning question is the Hobbesian question of order: What holds society together? In psychology the burning question is: What makes individuals behave and think the way they do? In anthropology the burning question is the nature of culture, its emergence and significance. In political science the burning question is how power is distributed and decisions are made in society. In economics the burning question is the production and distribution of wealth and resources in society. In education the burning question is how to transfer knowledge and teach people to learn. In evaluation the burning question is the utilization of information for decision making. Evaluation's burning question makes it a central discipline in the information age.

Standards for Evaluation

Evaluation as a profession has emerged since the 1960's as part of the legacy of the Great Society and War on Poverty programs. No sooner were those programs funded than questions began to be raised about how to monitor them, how to guarantee their implementation appropriately, and how to access their signifigance and impact. No sooner did evaluators begin working on these problems than the question was raised: How can evaluation be evaluated?

Evaluators were not altogether enthusiastic about the idea of evaluating evaluation. They explained that evaluations are very process-oriented, and therefore evaluation could not be assessed fairly by attention to limited a set of outcomes. They noted that there are multiple purposes and goals in an evaluation, and so it would not be possible to identify and measure only a limited set of evaluation objectives. They explained that the impact of evaluation was felt over the long-term, and therefore any evaluation that focused on immediate outcomes would miss much of the significant impact of evaluation. Indeed, it rapidly became clear that evaluators had learned their lessons (and excuses) from program staff with great astuteness.

Nevertheless, evaluators could not avoid evaluation. Thus, a national committee was assembled of eminent evaluators to develop standards for evaluation. Those standards were published in 1981 and represent criteria by which to judge the effectiveness of evaluation. With the official publication of new standadards of excellence, evaluation can be separated clearly from scientific research, and judged by different criteria. In the past an evaluation was considered "good" if the measuring instruments were carefully constructed, if samples were randomly drawn, and if results were statistically analyzed. Under the new standards, evaluations must also be useful, understandable, relevant, and practical.

This comprehensive effort at developing standards was hammered out over five years by a 17-member committee appointed by 12 professional organizations, with input from hundreds of practicing evaluation professionals. The standards published by the Joint Committee on Standards for Educational Evaluation (1981) dramatically reflect the ways in which the practice of evaluation has developed and changed during the last decade. Just prior to publication, Dan Stufflebeam, Chair of the Committee, summarized the committee's work as follows:

> The standards that will be published essentially call for evaluations that have four features. These are utility, feasibility, propriety, and accuracy. And I think it is interesting that the Joint Committee decided on that particular order. Their rationale is that an evaluation should not be done at all if there is no prospect for its being useful to some audience. Second, it should not be done if it is not feasible to conduct it in political terms, or practicality terms, or cost effectiveness terms. Third, they do not think it should be done if we cannot demonstrate that it will be conducted fairly and ethically. Finally, if we can demonstrate that an evaluation will have utility, will be feasible and will be proper in its conduct then they said we would turn to the difficult matters of the technical adequacy of the evaluation, and they have included an extensive set of standards in this area (Stufflebeam, 1980:90).

For each of these four basic concerns the Joint Committee articulated specific standards. Eight utility standards are identified to ensure that evaluations will serve the practical information needs of given audiences. These standards call for clear identification of audiences, writing clear and understandable reports, getting evaluations done on time, stating evaluator qualifications and biases, and taking responsibility for how an evaluation is used. Three feasibility standards are identified mandating that an evaluation should be realistic, diplomatic, and frugal. Eight propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results. Finally, eleven accuracy standards deal with such issues as the technical adequacy of evaluative information, sources of data, validity, reliability, data control, use of statistics, analysis of

qualitative information, drawing conclusions, and objectivity. A parallel set of standards have been developed by the Evaluation Research Society. Both sets of standards place primary emphasis on doing evaluations that are useful, practical, ethical, and accurate.

Barriers to Utilization

There are several barriers to the utilization of evaluation information for program improvement and decision making. I shall review three major barriers: fear, methods madness, and dogmatism.

Fear is quite natural to evaluation. Evaluation involves judgment -- judgment about good and bad, right and wrong -- and judgment, or the threat of judgment, can be frightening. There is fear of being judged, but more specifically, fear of being judged unfairly or inappropriately. There are related fears of data abuse and information misuse. These are legitimate fears based upon the real experiences of some victims of evaluation.

It is important to understand these fears and so I want to share with you a scholarly description of what the initial encounter between a program evaluator and program staff can be like. This scholarly analysis was conducted by none other than the eminent Dr. Seuss. In his story "What was I Scared of?" he tells of an encounter with "a pair of pale green pants with nobody inside them." The encounter immediately leads to a question: "What could those pants be there for?" (What could evaluation be there for?)

Dr. Seuss goes on to describe a variety of unexpected encounters with the empty pants and all kinds of attempts to avoid those encounters. But the attempts of avoidance eventually fail and the story culminates in a direct confrontation with the empty pants.

> I yelled for help. I screamed. I shrieked.
> I howled. I yowled. I cried,

"Oh, save me from these pale green pants
With nobody inside!"

Evaluation can be every bit as frightening as the pale green pants with nobody inside. But the Dr. Seuss story ends happily and it turns out that the empty pants are really quite friendly.

And, now, we meet quite often,
Those empty pants and I,
And we never shake or tremble.
We both smile
And we say
"Hi!

It would be my hope that while you may not find yourself in a torrid love affair with evaluation, you might at least find yourself on friendly, speaking terms with it.

In my work with program staff I find it is helpful to deal directly with the problem of fear. One way of doing that is to place it in a cultural context. (At this point the speaker puts on a somewhat frightening and awesome feather mask.) The mask I am wearing is from Carnival in Trinidad. Carnival is a time when people can disguise themselves and take on roles and behaviors that are different from their everyday routine. They can look at the world in new ways and, because of the masquerade, the world looks at them in new ways. Carnival is an integral part of Trinidadian culture. I like to wear the mask at the beginning of an evaluation process to confirm the staff's worst fears. The mask brings up images of voodoo, witch-doctoring, and evil. I then suggest that how one views the mask is very much a function of one's cultural perspective. I invite the staff to view evaluation in cultural terms. Evaluation is a subculture of Science. From a cultural perspective evaluation represents ways of thinking and looking at the world that are not typical of our everyday lives. Evaluation has its own language, its own perspective, its own rules and procedures, its own hierarchy, its own norms and values, and in total, its own

world view. In approaching a new culture we have been taught that cross-cultural sensitivity involves a suspension of judgment. An encouter or experience with a new culture seems to work best if we are tolerant, sensitive to different ways of doing things, and able to focus on understanding rather than judgment. I would invite you to approach evaluation in this way by attempting first to understand the values, world view, and perspective of evaluation before judging it. Evaluation, like a new culture, is not intrinsically good or bad, right or wrong. It simply is. It is one among many possible ways of looking at the world, and one of many possible ways of looking at remedial education programs. It is not the only way of looking at remedial education programs, but it can be an interesting and useful way of seeing what is happening in those programs. I would invite you, then, to experience the culture of evaluation and to experience what an evaluation perspective has to offer.

As you experience the culture of evaluation and get to be on speaking terms with it, it may come to be less frightening than it initially appears to be. From one perspective this mask initially seems quite frightening, but from another perspective it is quite beautiful. Like so many things it is a matter of how you look at it. (At this point the mask was removed.)

Another barrier to the utilization of evaluation is methods madness. Methods madness occurs when attention to methods and measurement so completely dominates the evaluation process that all else becomes secondary. The methods become an end in themselves rather than a means to the end of producing useful information. Methods take on a certain virtue that is intrinsic, a kind of rightness where using the proper scientific technique is more important than the quality and utility of the information generated by that technique.

Much methods madness stems from the longstanding debate over qualitative

versus quantitative methods. There is a status hierarchy in science which values some methods over others, and accords status to the practitioners of "hard" science over those who are "soft"-headed.

Dr. Seuss has described the confrontation of methods madness in his story of "The Sneetches."

> Now, the Star-Belly sneetches
> had bellies with stars.
> The Plain-Belly sneetches
> had none upon thars.

Dr. Seuss explains how the star-belly sneetches would have nothing to do with the plain-belly sneetches. They didn't socialize together or even talk to each other. In many social science departments the rife between quantitative and qualitative methodologists has been no less severe than the rife between star-belly and plain-belly sneetches.

The status hierarchy is upset in the story of the sneetches when a character shows up who is able to give the plain-belly sneetches stars by putting them through his very peculiar machine. The original star-belly sneetches are distraught when they find out that the plain-belly sneetches now have stars. This acquisition of new status by plain-belly sneetches is symbolic of the gradual status attained by qualitative methodologists as they developed their own journals, began to find publishing outlets for their monographs, and formed professional associations that gained attention and some degree of legitimacy for qualitative approaches.

The situation with the sneetches becomes further confused when the original star-belly sneetches are allowed to use the peculiar machine to remove their stars so they can again distinguish themselves from their disreputable counterparts. This is symbolic of the conversion experiences of leading quantitative methodologists like Campbell, Cronbach, and Guba who began writing

and giving speeches on the virtues of qualitative methods in the 70's. The original qualitative methodologists greeted these conversions with mixed feelings because their territory was being taken over by those who had formerly been their adversaries. At the same time some qualitative methodologists began touting the virtues of quantitative measurement. This created a period of great confusion for graduate students when it wasn't clear any longer which methods were "right." Qualitative methods, in the late 70's, became quite faddish while there were still single-minded advocates of quantitative measurement. Indeed, in the late 70's the field of evaluation was in considerable methodological confusion.

But Dr. Seuss does not like unhappy endings.

> I'm quite happy to say
> That the Sneetches got really quite smart on that day,
> The day they decided that Sneetches are Sneetches
> And no kind of Sneetch is the best on the beaches.
> That day, all the Sneetches forgot about stars
> And whether they had one, or not, upon thars.

In the 80's there has emerged in evaluation a consensus that evaluators are ill-advised to stake out a strong allegiance to any single methodological paradigm. There has emerged a consensus that methods are not right or wrong in absolute terms but rather methods must be judged in context, in terms of what questions are being asked and what kinds of data are needed for particular problems. Evaluation has come to be characterized by methodological tolerance where the issue is no longer routinely applying experimental designs and gathering quantitative data, but rather the question has become how to find methods that are appropriate to particular problems and situations. Multiple methods and combinations of qualitative and quantitative approaches have emerged as the approaches of choice in contrast to the former debate about the rightness or wrongness of quantitative versus qualitative methods. The madness of the earlier methods debates has given way to a focus on figuring out the important

questions and ways of addressing those questions with meaningful data. Methods are now used in the service of evaluation rather than being the masters of evaluators.

The third barrier to utilization is dogmatism. I have already discussed the falling away of dogmatism with regard to methods. But dogmatism threatens evaluation in other ways. The dogmatism of evalutation models and the dogmatism of preconceptions about what evaluation means can also raise barriers to the utilization of evaluative information. Here again Dr. Seuss has provided us with dramatic insights into the consequences of dogmatism.

Dr. Seuss describes the confrontation between a North-Going Zax and a South-Going Zax. When they meet, one going north and the other going south, they refuse to budge. Each insists that the other step aside, and each insists that he can outlast the other.

> I'll stay here, not budging! I can and I will if it makes you and me and the world stand still!

The Zax refused to budge but the world does not stand still. The world goes on, a new highway is built:

> And they built it right over those two stubborn Zax and left them there, standing, unbudged in their tracks.

People who approach evaluation dogmatically risk being left standing in their tracks as the world passes them by. The current practice of evaluation requires creativity, flexibility, situational responsiveness, and sensitivity to ways of raising questions that are particularly salient in specific contexts.

Moving Information Through a System

I now want to turn away from the barriers to utilization and look at some positive ways of moving information through a system. I'd like to illustrate some principles about moving information through a system by having you engage

in a quick exercise. I'd like to have you stand and join hands. There should be an endpoint and a starting point in the continuous chain of hands. Now I'll ask the person who is the starting point to squeeze the hand of the person next to him and for you to move those squeezes through the chain until the person at the end has received the squeeze. (The two sides of the audience engage in this exercise twice amidst much laughing, discussion, and clapping when one side achieves the end of the chain before the other.)

I would make five points about this exercise.

1. You can improve your ability to move information through the system with practice. The time it took to move the squeezes through the system on your first try was almost twice what it took the second time.

2. It takes time and effort to move information through the system. While time and effort can be reduced with practice and experience, time cannot be reduced to zero. There is a lower limit to the amount of time one can spend and still have something happen.

3. As information moves through a system it will wind around different parts of the system. Information use is not a linear process. It is not a straight line. It weaves around, sometimes in ways that can be anticipated, and many times in ways that are unanticipated.

4. Different people will receive and send evaluation messages in different ways. Some of you squeezed hard while others squeezed soft. Some like hard squeezes and others like soft squeezes. Some of you experienced the squeezes with some pain. I heard a few ouches as the information moved along. Passing information through the system can be painful. It will be received and passed on in different ways because people in the system are different.

5. Moving information through a system can be and should be fun.

An Exemplar of Utilization

Evaluators have started paying a lot of attention to the issue of utilization. How is it, they've asked themselves,that one can do research and evaluation work that people pay attention to and that makes a difference? Recent meetings of Evaluation Research Society and the Evaluation Network have been almost dominated by the question of utilization. Evaluators have both a personal commitment to doing useful work and a professional interest in it.

That's how I came into the field myself-I became interested in how people use information and how to increase that use. In so doing, I began to look for models and exemplars of evaluations that actually made a difference, and did a lot of work trying to find the factors that lead to utilization. I finally did find an exemplar and I want to share it with you. I want to go back several centuries to what is (as near as I can tell) history's first evaluation. It turns out that the first time around they did a pretty reasonable job of evaluation. I think there are some things that we can learn from history's first evaluation. The evaluation is recorded in the Old Testament in the first chapter of Daniel. You may recall that the Old Testament is largely the story of the children of Israel being in and out of captivity, and this is one of those periods of time when they were in captivity in Babylon. Basically, King Nebuchadnezzar had the idea that the Israelite children were culturally disadvantaged and he set up history's first remedial education program to overturn these cultural disadvantages so that they might be socialized into the dominant culture. He also had the vision, which reappears periodically in current times, of creaming his program so that he only took the better students. Indeed, this was a remedial program for the gifted!

As part of this special three-year remedial education program, Daniel and the other selected children were to be given a daily provision of meat and wine from the royal table. But rather than defile himself with the king's meat and wine, Daniel begged Ashpenaz, the chief eunuch (i.e., the program administrator), to put him and three other Israelite boys on a vegetarian diet. Ashpenaz, however, was worried that if they changed the diet the king would get upset because their "countenances" would not appear as fair as the other students. Here is a truly classic evaluation situation. The program administrator is afraid that if the students don't behave properly he'll lose

his head -- in this case, quite literally! Loss of one's head is usually a more figurative concern of program administrators (such as school principals and superintendents): if the clients (the students) don't behave properly and achieve according to what the king or the school board or the governor or someone else expects, then they're in big trouble.

So Ashpenaz went to Daniel and said he was worried about this whole situation. Couldn't they get together on something? So Daniel made a proposition that they conduct history's first evaluation: "Let us try it out. Let us change the diet, investigate it, evaluate the results, and--if we don't look worse after the experiment period--then we get to continue our vegetarian diet." Ashpenaz consented to a ten-day trial period. At the end of that period, Daniel and the other three boys appeared to Ashpenaz as even healthier then the other Israeli students. Those four boys were allowed to continue with the special diet for the entire three-year program. In fact, the students in the experimental program maintained their short-term gains over the long term, and ended at the top of their class!

I would submit to you that not only have we probably not had a more effective evaluation effort since that time, but there may not have been a more effective remedial educational program either!

Quite seriously, though, let us look at some of the factors which contributed to the high level of untilization of this particular evaluation exemplar: appropriate and relevant methods; the active involvement of the people who are going to be affected (including those who have to make decisions); a clear and limited focus to the research; and clear action implications specified in advance. These factors, the research on utilization shows, can make a critical difference in evaluation. This exemplar is very different from the way most evaluations are done.

This exemplar holds out the hope that evaluation can make a difference. It can be a tool for program improvement. At stake in applying evaluation is the quality of education for students in programs at all levels of achievement, including remedial and developmental education programs.

REFERENCES

The Joint Committe on Standards for Educational Evaluation, Standards for Evaluations of Educational Programs, Projects, and Materials. (New York, McGraw-Hill, 1981).

Patton, Michael Q., Utilization-Focused Evaluation. Beverly Hills, CA: Sage, 1978.

Patton, Michael Q., Qualitative Evaluation Methods. Beverly Hills, CA: Sage, 1980.

Patton, Michael Q., Creative Evaluation. Beverly Hills, CA: Sage, 1981.

Patton, Michael Q., Practical Evaluation. Beverly Hills, CA: Sage, 1982.

Seuss, Dr., The Sneeches and Other Stories. New York: Random House, 1961.

Stufflebeam, Daniel, L., "An Interview with Daniel L. Stufflebeam," Educational Evaluation and Policy Analysis, II (No. 2, 1980), 90; emphasis in the orginal.